

1.6.

(a) Consider

$$\bar{\ell} - H(\mathbf{p}) = \sum_{j=1}^D p_j (-\log q_j) + \sum_{j=1}^D p_j \log p_j = \sum_{j=1}^D p_j \log(p_j/q_j) .$$

Now extend \mathbf{p} and \mathbf{q} to one more alternative by defining $p_{D+1} = 0$ and $q_{D+1} = 1 - Q$. Now \mathbf{q} is a normalized probability distribution, i.e.,

$$\sum_{j=1}^{D+1} q_j = 1 .$$

(Notice that we do not require that q_{D+1} have the form 2^{-n} .) Now we have,

$$\sum_{j=1}^D p_j \ell_j - H(\mathbf{p}) = H(\mathbf{p}||\mathbf{q}) \geq 0 ,$$

which gives the desired result,

$$\bar{\ell} \geq H(\mathbf{p}) .$$

The conditions for equality in the relative entropy tell us that this inequality is saturated if and only if $p_j = q_j = 2^{-\ell_j}$ and $0 = p_{D+1} = q_{D+1} = 1 - Q$, i.e., $Q = 1$. Thus we have equality if and only if the probabilities are all negative powers of 2; in this situation, we can construct code words from a complete tree with $\ell_j = -\log p_j$.

(b) We make an obvious choice by requesting codeword lengths $\ell_j = -\lceil \log p_j \rceil$, where $\lceil x \rceil$ is the smallest integer $\geq x$. We have

$$\log p_j \leq \ell_j < -\log p_j + 1 .$$

Using the left inequality, we have $q_j = 2^{-\ell_j} \leq p_j$, which gives

$$Q = \sum_{j=1}^D q_j \leq \sum_{j=1}^D p_j = 1 .$$

The Kraft inequality thus assures us that there are code words of the requested length. This is called a *Shannon-Fano code*. Using the right inequality, we have

$$\bar{\ell} < \sum_{j=1}^D p_j (-\log p_j + 1) = H(\mathbf{p}) + 1 .$$

(c)

1. Let's order the alternatives from most likely to least likely, i.e., $p_1 \geq p_2 \geq \dots p_D$.

Now we show the following: *any optimal code can be converted to an equivalent optimal code in which the two least likely alternatives, x_{D-1} and x_D , have code words of the same (maximum) length, these two code words differing only in the final letter. The code word for x_D ends in 0, and the code word for x_{D-1} ends in 1.*

If some alternative has a code word longer than the code word of x_D , we can simply exchange the two code words, not increasing the average length in the process. Thus, if the code is optimal, the least likely alternative must have a code word of the longest length in the code.

If x_D has a code word such that flipping the last letter is not a code word, then we can shorten its code word by deleting the final letter, while maintaining the prefix-free condition. Thus, if the code is optimal, x_D must have a *partner* whose code word differs only in the last letter.

If x_{D-1} has a shorter code word than the partner of x_D , we can decrease—really, not increase—the average length by exchanging the two code words. Thus, in an optimal code, the two least likely alternatives have code words of the same (maximum) length. Now, if x_{D-1} is still not the partner of x_D , because it has the same code word length as the partner, we can exchange the code words of x_{D-1} and the partner without changing the average length. We can choose the code word of x_D to end in 0 and that of x_{D-1} to end in 1 without changing the average length.

We have shown what we intended.

2. Now amalgamate x_{D-1} and x_D into a single alternative x'_{D-1} , with probability $p'_{D-1} = p_{D-1} + p_D$ and with code word obtained by lopping off the final letter of the code word of x_D (or x_{D-1}). Define a new set of (primed) alternatives that consists of x'_{D-1} and all the original alternatives up through $D - 2$. The average length of the original code words is

$$\begin{aligned}
 \bar{\ell} &= \sum_{j=1}^D p_j \ell_j \\
 &= \sum_{j=1}^{D-2} p'_j \ell'_j + \underbrace{(p_{D-1} + p_D)}_{= p'_{D-1}} \times \underbrace{\ell_D}_{= \ell'_{D-1} + 1} \\
 &= \sum_{j=1}^{D-1} p'_j \ell'_j + p_{D-1} + p_D \\
 &= \bar{\ell}' + p_{D-1} + p_D .
 \end{aligned}$$

Thus the problem of finding an optimal code for the original alternatives is reduced to finding an optimal code for the primed alternatives.

Now we just repeat step 1, and we have shown that any optimal code can be converted to a code constructed by the Huffman procedure.